Causality and Experiments

Michael R. Roberts

Department of Finance
The Wharton School
University of Pennsylvania

April 13, 2009

Motivation

- Do hospitals make people healthier? (Causation)
- Compare avg health of hospital visitors no non-visitors (2005 NHIS)
  1. Mean health status of hospital visitors = 3.21
  2. Mean health status of hospital non-visitors = 3.93**

  Hospitals make people *less* healthy. (Hospital can be dangerous.)
- Or, hospital visitors – who self-select – are different from non-visitors in a way that is correlated with health.
- **Non-Random Selection** is a major obstacle in empirical work.
- Goal here is to develop a simple framework in which we can understand the problem and identify ways to address it.

Notation: Potential Outcomes

- Treatment (e.g., go to hospital) indicator: $D_i = \{0, 1\}$
- Outcome variable (e.g., health status): Y_i
- Question: Is Y_i *affected* by treatment?
- Setup: There are *two* **Potential Outcomes** for each individual i,

$$\textit{Potential Outcome} = \begin{cases} Y_{1i} & \text{if } D_i = 1 (\text{i.e., receive treatment}) \\ Y_{0i} & \text{if } D_i = 0 (\text{i.e., receive treatment}) \end{cases}$$

- Answer: For each person i we want to know the difference $Y_{1i} - Y_{0i}$ This is causal effect of treatment on individual i.
- Problem: For each person i, we only observe one of the outcomes absent being able to rewind the clock and change treatment status for a person.
  - Unobserved outcome is **counterfactual**

Notation: Observed Outcomes

- The **Observed Outcome** is Y_i
- Observed Outcome can be written in terms of Potential Outcomes:

$$\begin{aligned} Y_i &= \begin{cases} Y_{1i} & \text{if } D_i = 1 \text{(i.e., receive treatment)} \\ Y_{0i} & \text{if } D_i = 0 \text{(i.e., receive treatment)} \end{cases} \\ Y_i &= Y_{0i} + (\underbrace{Y_{1i} - Y_{i0}}_{\text{Causal Effect}})D_i (= D_i Y_{1i} + (1 - D_i) Y_{0i}) \end{aligned}$$

- Note: Causal (a.k.a. Treatment) effect can be different for different people i
- Since we never observe Y_{1i} and Y_{0i} for the same person, we must infer treatment effect by comparing treated outcomes to untreated outcomes.

Treated Versus Untreated Comparison

- What is difference in expectations across treated and untreated?

$$\underbrace{E[Y_i|D_i=1]-E[Y_i|D_i=0]}_{\text{Observed Dif in Outcomes}} = E[Y_{1i}|D_i=1]-E[Y_{0i}|D_i=0]$$
$$= \underbrace{E[Y_{1i}|D_i=1]-E[Y_{0i}|D_i=1]}_{\text{Avg treatment effect on treated (ATT)}}$$
$$+ \underbrace{E[Y_{0i}|D_i=1]-E[Y_{0i}|D_i=0]}_{\text{Selection Bias}}$$

  1^{st} = from def of observed outcome in terms of potential outcomes.
  2^{nd} = comes from $\pm\ E[Y_{0i}|D_i=1]$ on RHS.
- Problem: Observed difference in outcomes adds *selection bias* term to the causal term we want
- Selection term = dif in avg Y_{0i} between the treated and untreated. E.g., sick more likely to visit hostpial $\implies$ worse Y_{0i} $\implies$ negative selection bias.

Random Assignment

- Random assignment overcomes selection bias because treatment status will be independent of potential outcomes
- Reconsider selection term under random assignment

$$E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] = E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 1] = 0$$

- Since outcomes are idenpendent of treatment stats, we can swap

$$E[Y_{0i}|D_i = 0] for E[Y_{0i}|D_i = 1]$$

- Reconsider the causal term under random assignment

$$E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] = E[Y_{1i} - Y_{0i}|D_i = 1] = E[Y_{1i} - Y_{0i}]$$

- **Random assignment eliminates selection bias.**

Labor Economics Example

- Evaluation of gov't-subsidized tranining programs. Do they increase employment and earnings?
- Compare earnings after training of participants to nonparticipants and trainees earn less than plausible comparison groups (e.g., Ashenfelter 1978, Ashenfelter and Card (1985), LaLonde (1995)).
- Selection bias: training programs serve people with low-earnings potential so $E[Y_{0i}|D_i = 1] < E[Y_{0i}|D_i = 0] \implies$ negative selection bias

$$E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] < 0$$

  leads to differences in observed avgs across groups that are biased downward.
- Randomized trials generate positive effects of training programs (Lalonde (1986) and Orr et al. (1996))

What Makes for a Good Randomized Experiment?

- Does randomization balance subject characteristics across treatment & control groups?
  - The two groups should have similar characteristics and outcomes pre-treatment.
- With randomization, we can estimate causal effect by comparing sample means and performing t-test.
- If worried about SEs, use regression framework and a dummy indicating treatment status

$$Y_i = \alpha + \beta D_i + \varepsilon_i$$

  can estimate cluster or heteroskedastic-robust SEs.
- To determine economic significance, compre estiamted affect to a measure of spread (e.g., standard deviation, interquartile range).
- The same suggestions apply to natural or quasi-natural experiments!

Regression Analysis of Experiments I

- Assume constant (homogenous) treatment effect across $i, \implies Y_{1i} - Y_{0i} = \rho \forall i$. In regression form:

$$Y_i = \underbrace{\alpha}_{E(Y_{0i})} + \underbrace{\rho}_{Y_{1i}-Y_{0i}} D_i + \underbrace{\eta_i}_{Y_{0i}-E(Y_{0i})}$$

- Where does this come from? Consider the potential outcomes:

$$\begin{aligned} Y_{1i} &= \alpha + \rho + \eta_i \text{ (when treated)} \\ Y_{0i} &= \alpha + \eta_i \text{ (when treated)} \end{aligned}$$

- Subtract Y_{0i} from Y_{1i} to get $\rho = Y_{1i} - Y_{0i}$
- Take *unconditional* expectation of Y_{0i} to get $\alpha = E[Y_{0i}]$

Regression Analysis of Experiments II

- Consider the conditional expectations of the regression equation:

$$\begin{aligned} E[Y_i|D_i = 1] &= \alpha + \rho + E[\eta_i|D_i = 1] \\ E[Y_i|D_i = 0] &= \alpha + E[\eta_i|D_i = 0] \end{aligned}$$

which implies the estimated treatment effect is

$$\begin{aligned} E[Y_i|D_i = 1] = E[Y_i|D_i = 0] &= \underbrace{\rho}_{\text{Treatment Effect}} \\ &+ \underbrace{E[\eta_i|D_i = 1] - E[\eta_i|D_i = 0])}_{\text{Selection Bias}} \end{aligned}$$

- Randomization $\implies$ selection bias $= 0$ since

$$E[\eta_i|D_i = 1] = E[\eta_i|D_i = 0]$$

Selection Bias = Nonzero-mean Conditional Error

- Last eqn on prev slide shows that:

Selection Bias = Nonzero-mean Conditional Error
= Correlation between Regressor (D_i) and Error (η_i)

- Recall from slide 5 that selection bias is:

$$(E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0])$$

- Combining with regression results means:

$$(E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]) = (E[\eta_i|D_i = 1] - E[\eta_i|D_i = 1])$$

- Nonzero-conditional mean error reflects the difference in (no-treatment) potential outcomes between the treated and untreated.
- In hospital example, treated had worse health in no-treatment state than untreated in no-treatment health state.
- Must have similar treatment and control groups *outside* treatment.

Heterogeneous Treatment Effects I

- What if $\rho = \rho_i \implies$ treatment effect varies across individuals?
- Regression model is now:

$$Y_i = \alpha_i + \rho_i D_i + \eta_i$$

- Taking conditional expectations:

$$\begin{aligned} E[Y_i|D_i = 1] &= \alpha + E[\rho_i|D_i = 1] + E[\eta_i|D_i = 1] \\ E[Y_i|D_i = 0] &= \alpha + E[\eta_i|D_i = 0] \end{aligned}$$

and subtracting equations

$$\begin{aligned} E[Y_i|D_i = 1] &- E[Y_i|D_i = 0] \\ &= \underbrace{E[\rho_i|D_i = 1]}_{\text{Avg. Treatment Effect of Treated (ATT)}} \\ &+ \underbrace{(E[\eta_i|D_i = 1] - E[\eta_i|D_i = 0])}_{\text{Selection Term}} \end{aligned}$$

Heterogeneous Treatment Effects II

- How do we recover the average treatment effect, $E[\rho_i]$?
- Express ATE ($E[\rho_i]$) in terms of ATT ($E[\rho_i|D_i = 1]$).

$$\begin{aligned} E[\rho_i] &= Pr(D_i = 0)E(\rho_i|D_i = 0) + Pr(D_i = 1)E(\rho_i|D_i = 1) \\ &= Pr(D_i = 0)E(\rho_i|D_i = 0) + (1 - Pr(D_i = 0))E(\rho_i|D_i = 1) \\ &= Pr(D_i = 0)\left[E(\rho_i|D_i = 0) - E(\rho_i|D_i = 1)\right] + E(\rho_i|D_i = 1) \end{aligned}$$

- Now plug into last eqn on prev slide to get:

$$\begin{aligned} E[Y_i|D_i = 1] &- E[Y_i|D_i = 0] \\ &= \underbrace{E[\rho_i]}_{\text{Avg. Treatment Effect}} + \underbrace{(E[\eta_i|D_i = 1] - E[\eta_i|D_i = 0])}_{\text{Selection Term}} \\ &+ Pr(D_i = 0)\underbrace{(E[\rho_i|D_i = 1] - E[\rho_i|D_i = 0])}_{\text{Heterogenous Treatment Effects}} \end{aligned}$$

- Extra term = difference in avg gains from treatment across groups
- Randomization solves both selection biases

Control Variables

- If you have a proper experiment, you shouldn't have to control for confounding influences, X.
  - In linear regression, controls don't matter. In nonlinear setting this is problematic (see Freedman (??))
  - In hospital example, may want to control for sex, race, past health, habits (e.g., smoker), etc. for each person.
- If Controls uncorrelated with treatment status, then estimated effect should be unaffected by their inclusion.
- Controls can generate more precise estimates by absorbing residual variation.

Big Picture

- Rarely do we have **randomized experiments**
- We have **observational studies** where non-random selection is key concern.
- Further, homogenous treatment effect is often a stretch.
- Goal is to overcome the selection bias (and deal with heterogeneous treatment effects) to make **causal statements**
  - Hospitals make people healthier
  - CEOs create value for firms
  - Acquisitions destroy value
  - Firms issue equity to take advantage of information-based mispricing.
  - Etc.
- Rest of program evaluation component of course focuses on how to overcome selection bias to make causal inferences
- More broadly, selection pops up in other contexts (e.g., structural estimation) so we must understand the problem and how to overcome it.